APPARATUS AND METHOD FOR NORMALIZING INPUT DATA OF ACOUSTIC MODEL AND SPEECH RECOGNITION APPARATUS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit under 35 USC 119(a) of Korean Patent Application No. 10-2015-0144947 filed on Oct. 16, 2015, in the Korean Intellectual Property Office, the entire disclosure of which is incorporated herein by reference for all purposes.

BACKGROUND

[0002] 1. Field

[0003] The following description relates to technology for normalizing input data of an acoustic model for gradual decoding in speech recognition.

[0004] 2. Description of Related Art

[0005] In general, a speech recognition engine consists of an acoustic model, a language model, and a decoder. The acoustic model calculates pronunciation-specific probabilities for each frame of an input speech signal, and the language model provides information on how frequently a specific word or sentence is used. The decoder calculates which word or sentence is similar to an input speech based on the information provided by the acoustic model and the language model, and outputs the calculation result. A Gaussian mixture model (GMM) acoustic model has been generally used, and speech recognition performance is improving lately with the advent of a deep neural network (DNN) acoustic model. A bidirectional recurrent deep neural network (BRDNN) calculates pronunciation-specific probabilities for each frame of a speech in consideration of bidirectional information, that is, preceding and subsequent frame information, and thus receives the speech as a whole. When each frame of a speech signal input during model training is represented as an N-dimensional vector, a BRDNN acoustic model performs normalization so that each dimensional value of the vector is within a specific range. While normalization may be generally performed based on whole training data or each utterance, the BRDNN acoustic model performs normalization in units of utterances.

SUMMARY

[0006] This summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used as an aid in determining the scope of the claimed subject matter.

[0007] In one general aspect, an apparatus for normalizing input data of an acoustic model includes a window extractor configured to extract windows of frame data to be input to the acoustic model from frame data of a speech to be recognized; and a normalizer configured to normalize the frame data to be input to the acoustic model in units of the extracted windows.

[0008] The window extractor may be further configured to consecutively extract the windows in units of a predetermined number of frames of the frame data of the speech to be recognized while the frame data of the speech to be recognized is being input.

[0009] The normalizer may be further configured to normalize frames belonging to a current window together with padding frames added to both sides of the current window.

[0010] The normalizer may be further configured to normalize frames belonging to a current window in consideration of frames belonging to preceding windows of the current window.

[0011] The normalizer may be further configured to normalize the frames belonging to the current window in consideration of the frames belonging to the preceding windows and frames of training data in response to a total number of the frames belonging to the current window and the frames belonging to the preceding windows being insufficient for speech recognition.

[0012] The normalizer may be further configured to acquire a number of frames corresponding to a difference between the total number of the frames and a reference value from the training data in response to the total number of the frames being less than the reference value.

[0013] The normalizer may be further configured to normalize the frame data belonging to the extracted windows so that the frame data belonging to the extracted windows has an average of 0 and a standard deviation of 1.

[0014] In another general aspect, a method of normalizing input data of an acoustic model includes extracting windows of frame data to be input to the acoustic model from frame data of a speech to be recognized; and normalizing the frame data to be input to the acoustic model in units of the extracted windows.

[0015] The extracting of the windows may include consecutively extracting the windows in units of a predetermined number of frames of the frame data of the speech to be recognized while the frame data of the speech to be recognized is being input.

[0016] The normalizing of the frame data may include normalizing frames belonging to a current window together with padding frames added to both sides of the current window.

[0017] The normalizing of the frame data may include normalizing frames belonging to a current window in consideration of frames belonging to preceding windows of the current window.

[0018] The normalizing of the frame data may include normalizing the frames belonging to the current window in consideration of the frames belonging to the preceding windows and frames of training data in response to a total number of the frames belonging to the current window and the frames belonging to the preceding windows being insufficient for speech recognition.

[0019] The normalizing of the frame data may include comparing the total number of the frames belonging to the current window and the preceding windows with a reference value in response to the current window being extracted; and acquiring a number of frames corresponding to a difference between the total number of the frames and the reference value from the training data in response to the total number of the frames being less than the reference value.

[0020] The normalizing of the frame data may include normalizing the frame data belonging to the extracted windows so that the frame data belonging to the extracted windows has an average of 0 and a standard deviation of 1.